Off-nominal event analysis in autonomous flights based on explainable artificial intelligence

Shivakumar Ranganathan | Hari Ilangovan | Newton Campbell | Michael Acheson | Irene Gregory

2024 AIAA SciTech

Orlando, FL

January 9, 2024

Outline

- Overview
- Background
- eXplainable AI (XAI)
- Case Studies
- Conclusion

Overview

- Mission of NASA Urban Air Mobility (UAM) Program: Perform autonomous flights

  —under all weather conditions

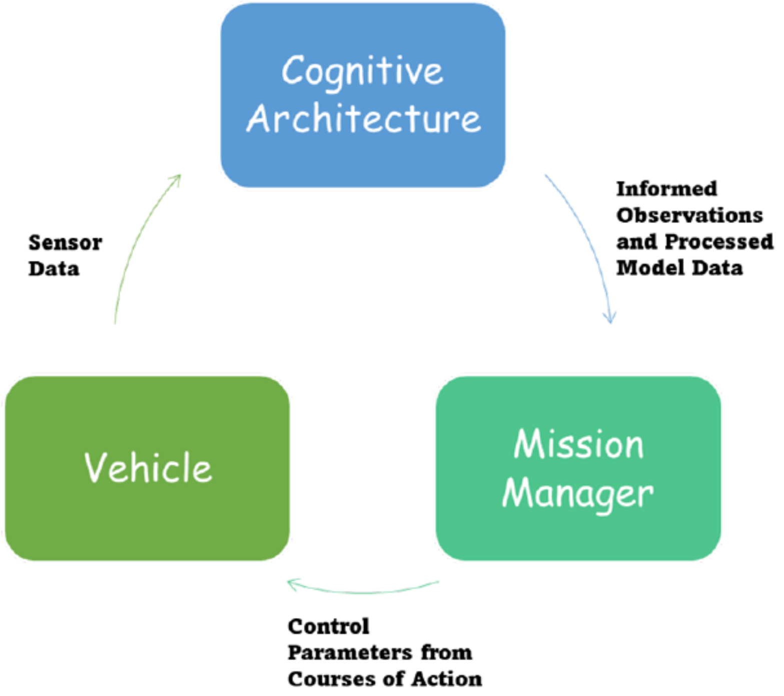
  —in a high-density airspace with complex urban environment

  —with guaranteed levels of safety

- A Cognitive Mission Manager (Cognitive Architecture + Mission Manager) develops and decides among Courses of Actions (CoAs) to enable the vehicle to complete the mission safely

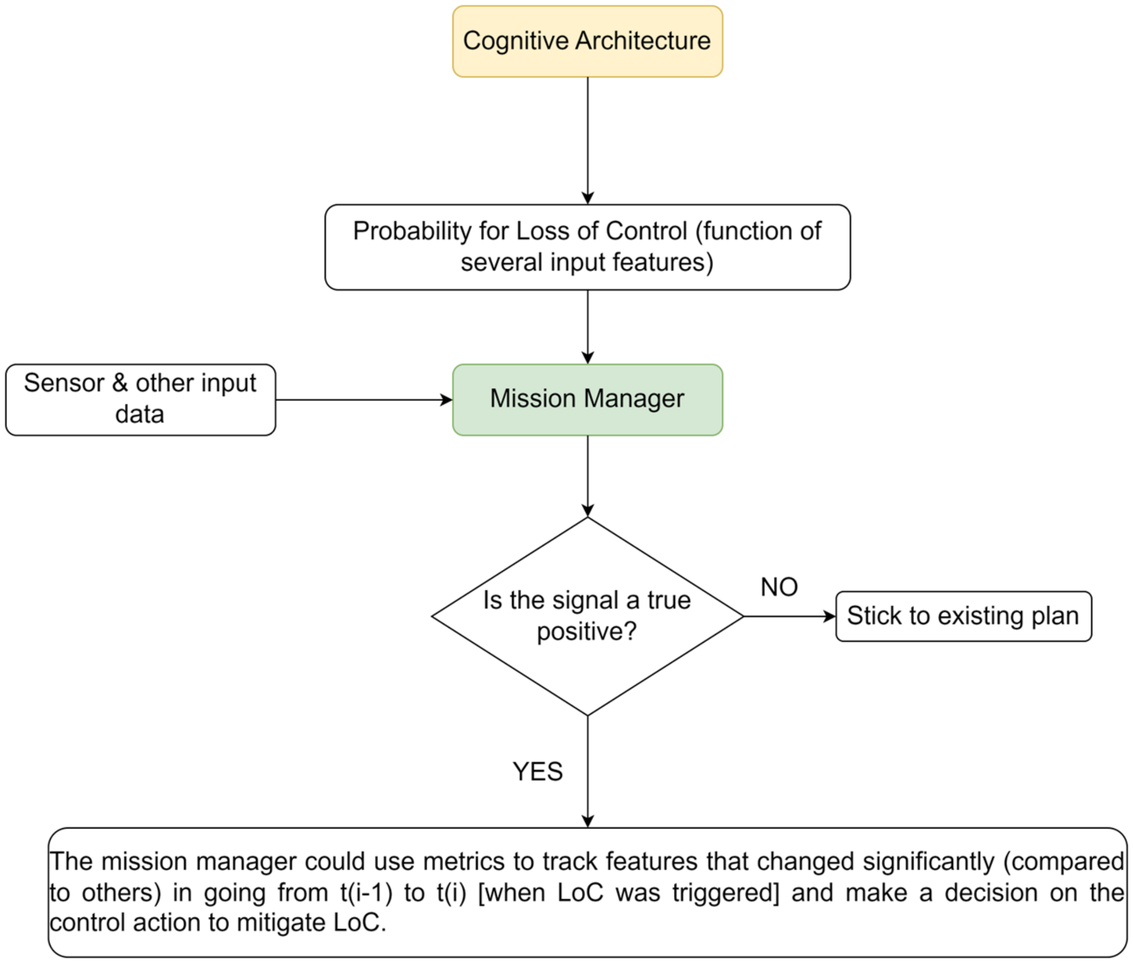
- Cognitive Architecture represents the Observe part of the OODA (Observe-Orient-Decide-Act) loop and hosts a variety of algorithms (Hankel Alternative View of Koopman, Conditional Variational Autoencoder) to assess the vehicle's current and future capabilities

- Mission Manager represents the Orient-Decide-Act parts and controls setting of short-term and long-term goals for the vehicle during flight, as well as CoAs



Overview

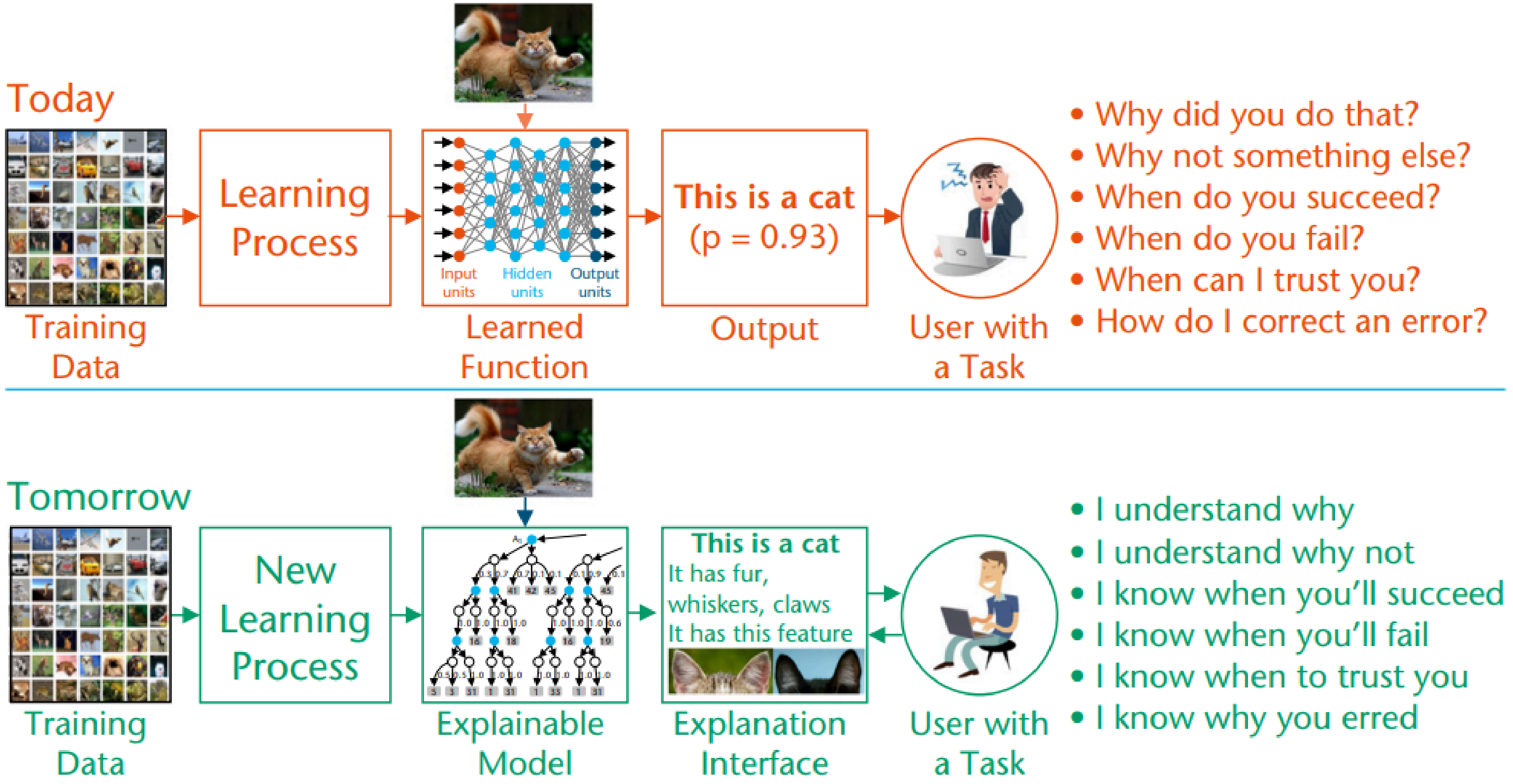
Information flow within the cognitive mission manager in a Loss of Control scenario



Which of the input features most likely triggered the off-nominal flag?

eXplainable AI (XAI)

- eXplainable AI (XAI)[1,2]: AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future



[1] Gunning, D and Aha D.W., 2019. DARPA's explainable artificial intelligence program. AI Mag, 40(2), p.44.D., 2022. Feature Importance for Time Series Data: Improving KernelSHAP. arXiv preprint arXiv:2210.02176.
[2] Gunning, D., Vorm, E., Wang, Y. and Turek, M., 2021. DARPA's explainable AI (XAI) program: A retrospective. Authorea.

eXplainable AI (XAI)

Why XAI?

- Transparency: Autonomous systems that can explain their decision-making processes provide greater transparency, which can help build trust between users and the system
- Accountability: In cases where an autonomous system causes harm, XAI can help understand why the system made the decision it did
- Regulation: Transparency in the decision-making process can help autonomous systems to comply with legal and ethical standards
- Human-AI collaboration: XAI can facilitate collaboration between humans and AI by providing a shared understanding of how the system arrived at a decision (would help improve the system over time)

eXplainable AI (XAI)

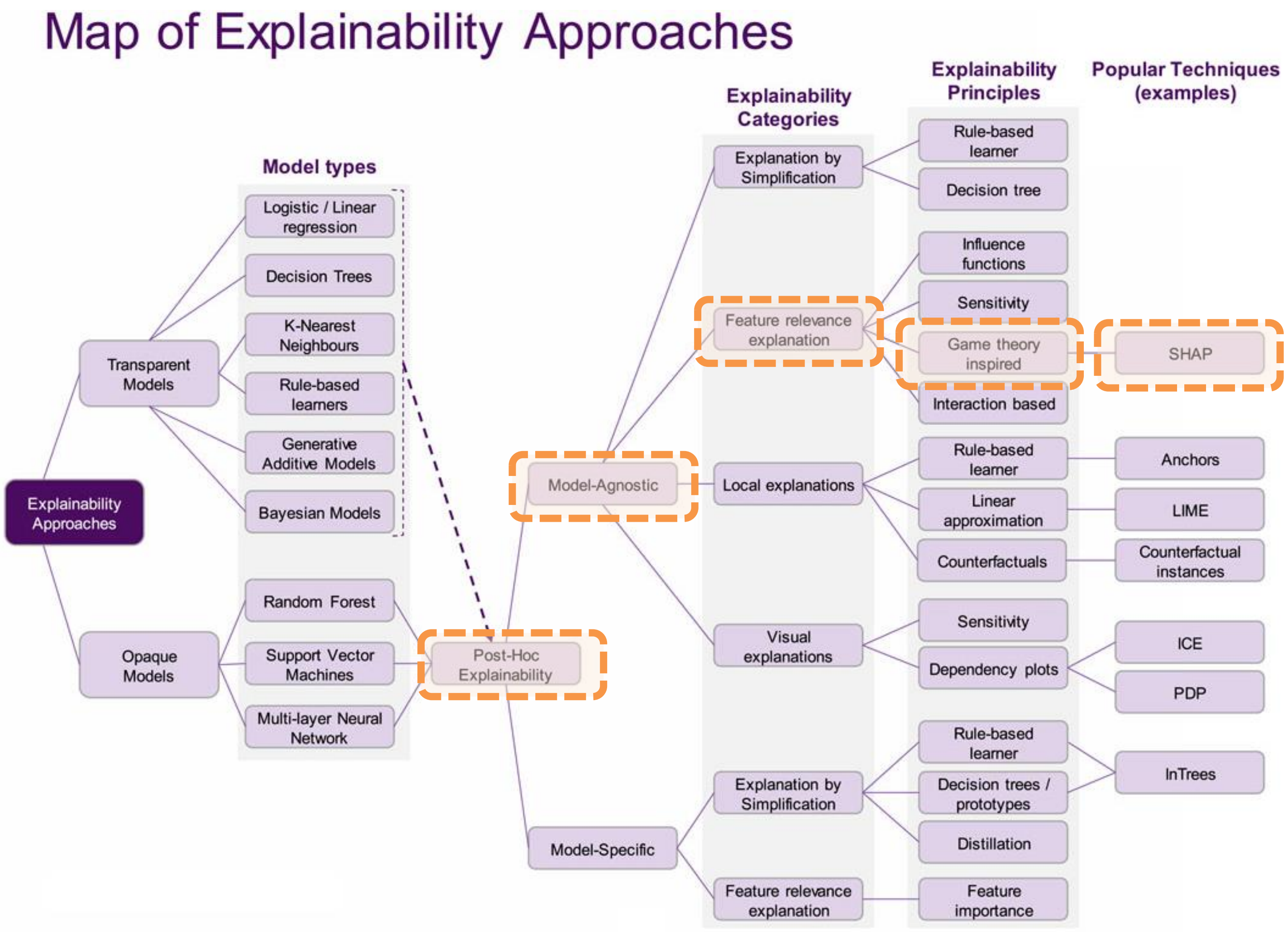
Prior Work:

- Rudin (2019)[3] presented a case for developing models that are interpretable in the first place rather than attempting to explain black box models
- Lundberg & Lee (2017)[4] used SHAP (SHapley Additive exPlanations) to alleviate the tension between accuracy and interpretability in ML models
- Schlegel et al (2019)[5] conducted preliminary experiments to assess the quality of XAI method explanations on a range of datasets and concluded that SHAP works robust for all models, but others like DeepLIFT, LRP, and Saliency Maps work better with specific architectures
- Villani et al (2022)[6] consider Shapley value based approaches to feature importance, applied in the context of time series data and demonstrate how feature importance extracted from this technique can perform "event detection"

[3]Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1 (5).
[4]Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
[5]Schlegel, U., Arnout, H., El-Assady, M., Oelke, D. and Keim, D.A., 2019, October. Towards a rigorous evaluation of XAI methods on time series. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 4197-4201). IEEE.
[6]Villani, M., Lockhart, J. and Magazzeni, D., 2022. Feature Importance for Time Series Data: Improving KernelSHAP. arXiv preprint arXiv:2210.02176.

eXplainable AI (XAI)

XAI Approaches:[7,8]



[7] Belle, V. and Papantonis, I., 2021. Principles and practice of explainable machine learning. Frontiers in big Data, p.39.
[8] https://www.nobelprize.org/prizes/economic-sciences/2012/shapley/facts/

eXplainable AI (XAI)

XAI Technique based on Shapley values:

- Satisfies the four Axioms[9] required to achieve a fair contribution
    a) Efficiency: The sum of the contributions of all features should give the total payout (prediction)

    b) Symmetry: Contributions of two feature values j and k should be the same if hey contribute equally to all possible coalitions

    c) Dummy: If a feature j does not change the prediction value, no matter to which coalition of feature values it is added, its Shapley value should be 0

    d) Additivity: Overall utility or performance of a set of features should be the sum of the utilities or performances of the individual features

- Model Agnostic: It can be used with any machine learning model and is not limited to any specific algorithm or architecture

- Generates local, global explanations and provides insight into the behavior of the model both at a specific instance and across the entire dataset

[9]Jakubiak, N., 2022. Analysis of explainable artificial intelligence on time series data.

eXplainable AI (XAI)

- Example: 3 players share a taxi. The costs for each individual journey:
  - Player 1: 6
  - Player 2: 12
  - Player 3: 42

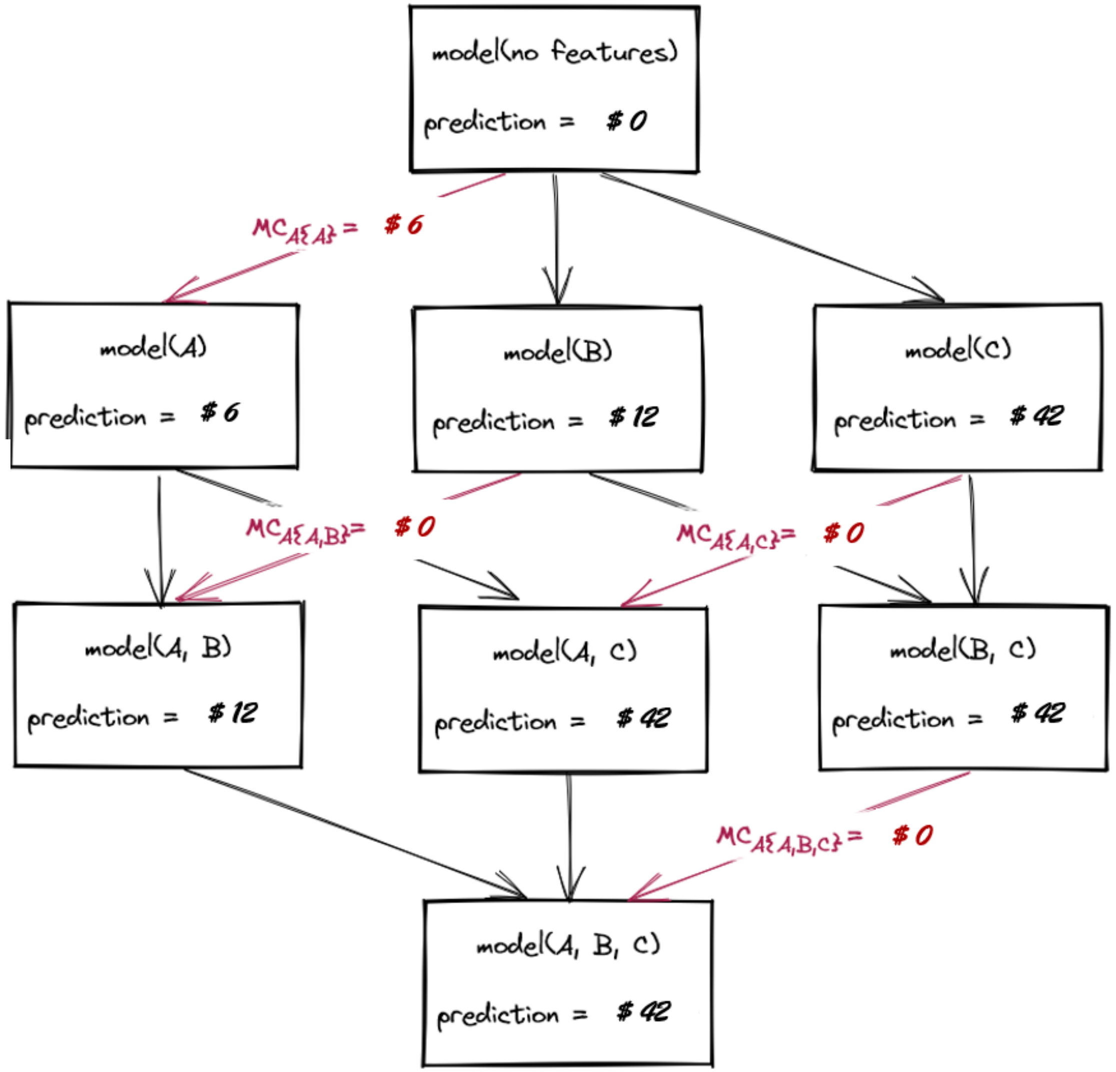
How much should each individual contribute?

$$shapley(Person\ 1) = 1/3\times 6 + 1/6\times 0 + 1/6\times 0 + 1/3\times 0 = 2$$

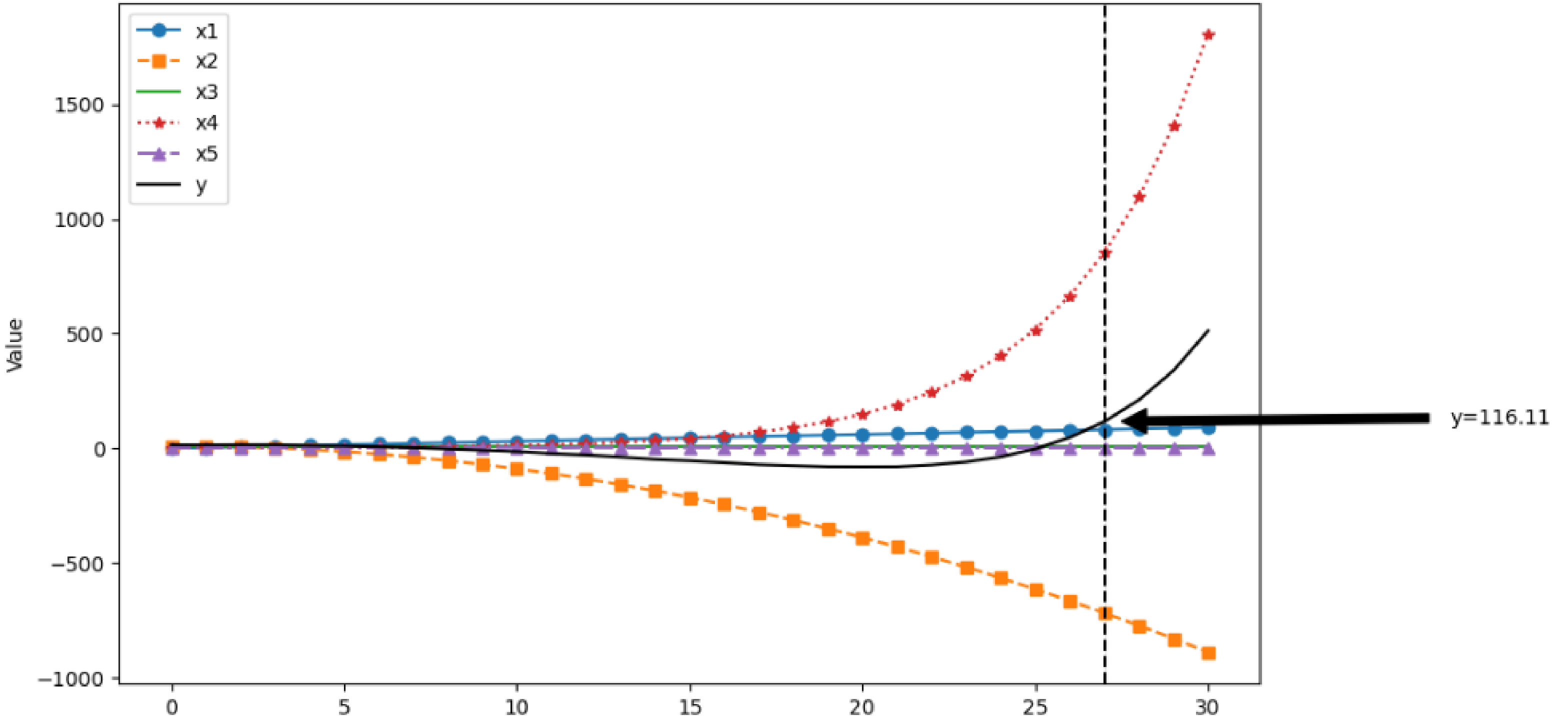
$$shapley(Person\ 2) = 1/3\times 12 + 1/6\times 6 + 1/6\times 0 + 1/3\times 0 = 5$$

$$shapley(Person\ 3) = 1/3\times 42 + 1/6\times 36 + 1/6\times 30 + 1/3\times 30 = 35$$

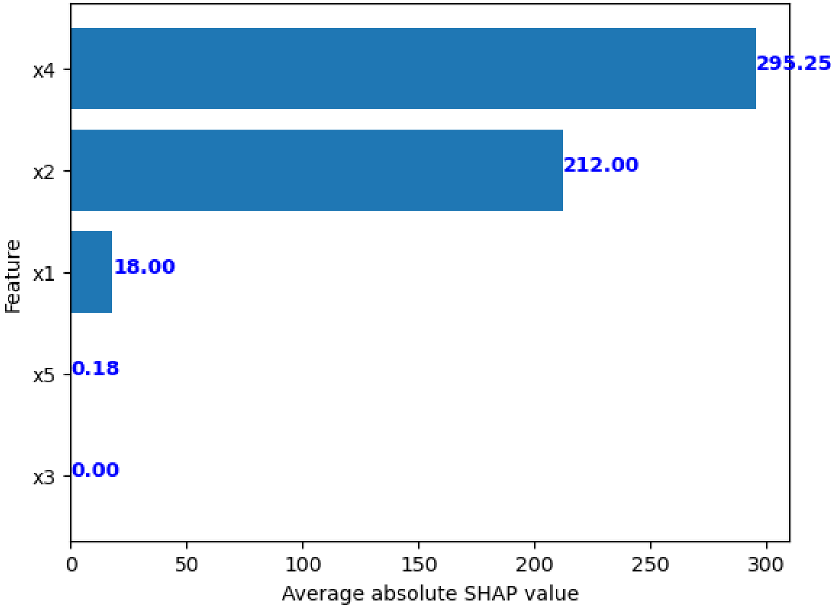
$$Total\ Fare = shapley(Person\ 1) + shapley(Person\ 2) + shapley(Person\ 3) = 42$$

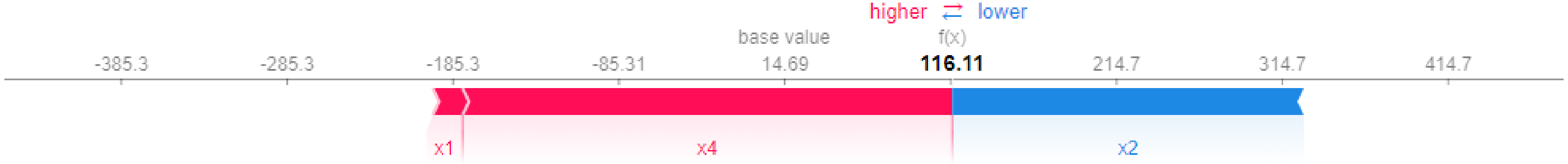


Case Study 1: Known dataset

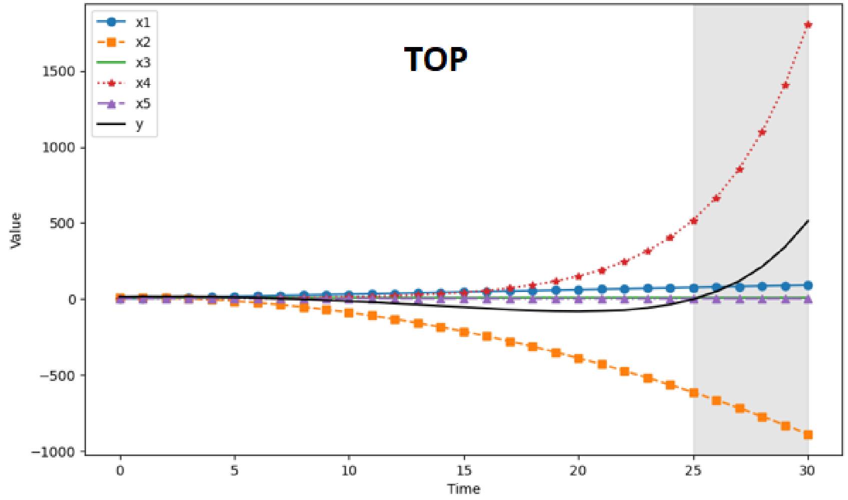


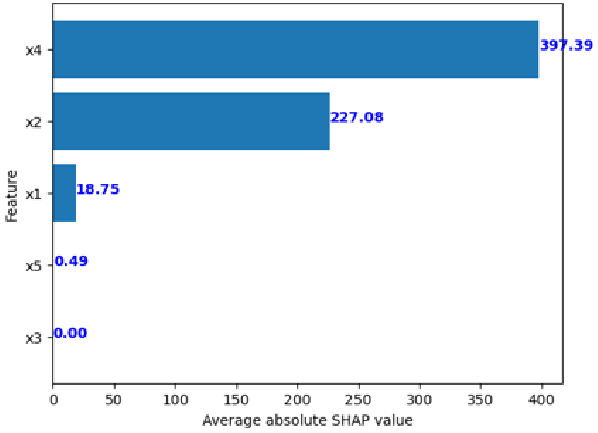

```
time instance:  27
Baseline:  14.68511625580915
SHAP(x1):  17.999999999999982
SHAP(x2):  -212.0
SHAP(x3):  0.0
SHAP(x4):  295.25037440896205
SHAP(x5):  0.17711548480100703
sum: 116.11260614957219
y_instance: 116.11260614957217
```

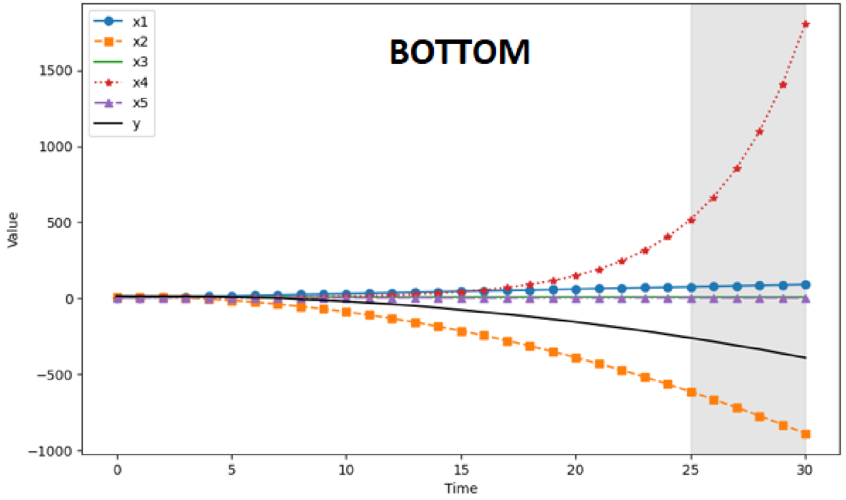


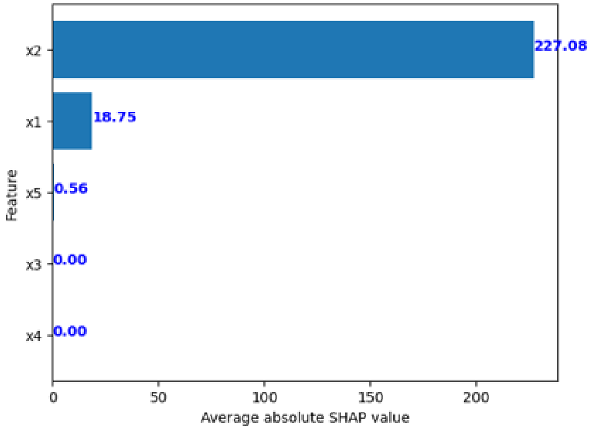


Case Study 1: Known dataset



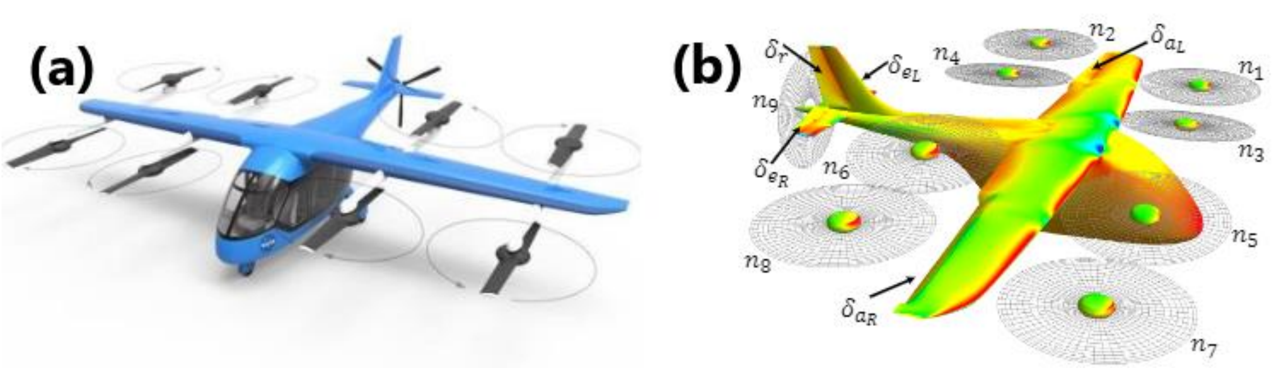






Effect of specific input feature ($x4$) on the target variable in the time interval $25 \leq t \leq 30$ (Consistent with the nature of data, top plot captures dependence of target on ($x4$), whereas the bottom plot does not)

Case Study 2: Analysis of nominal flight data

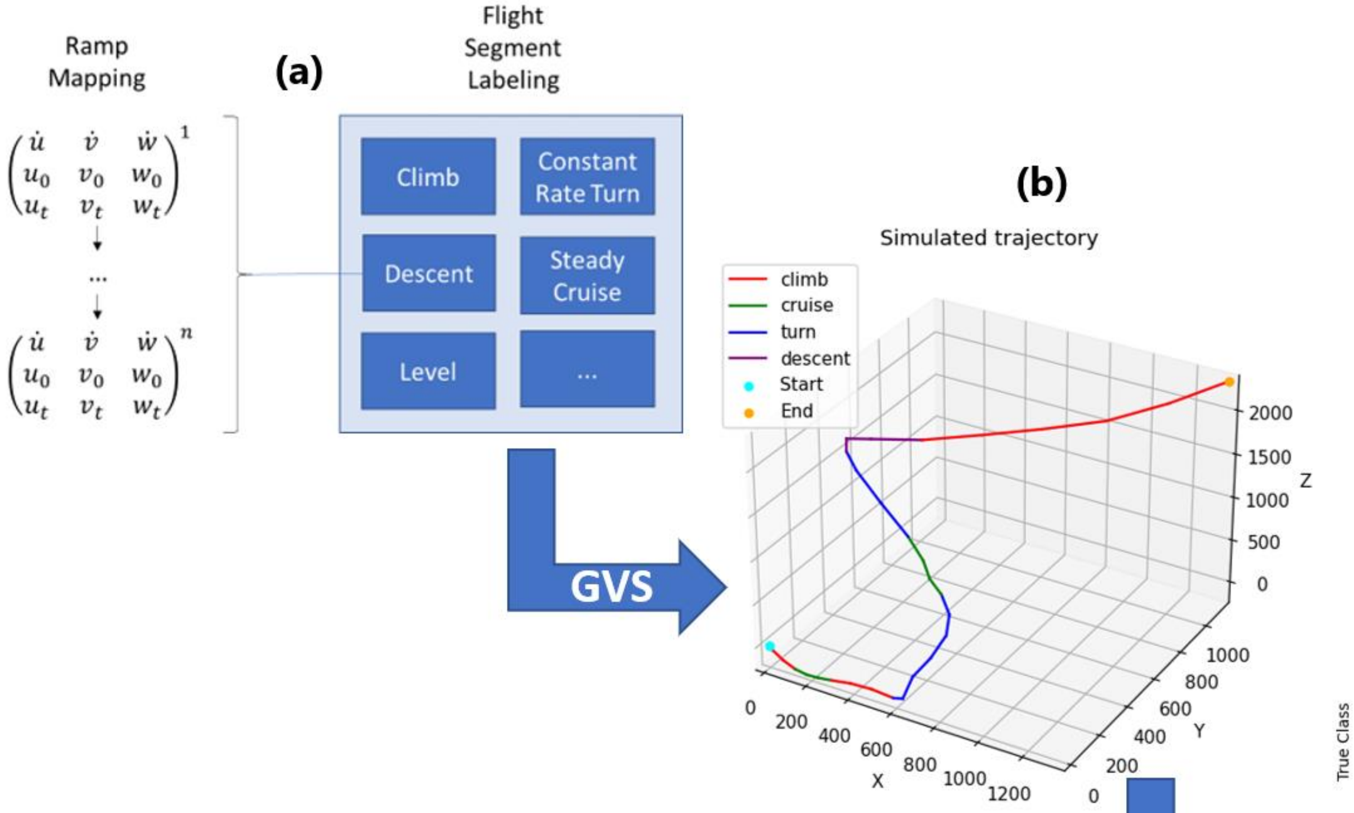


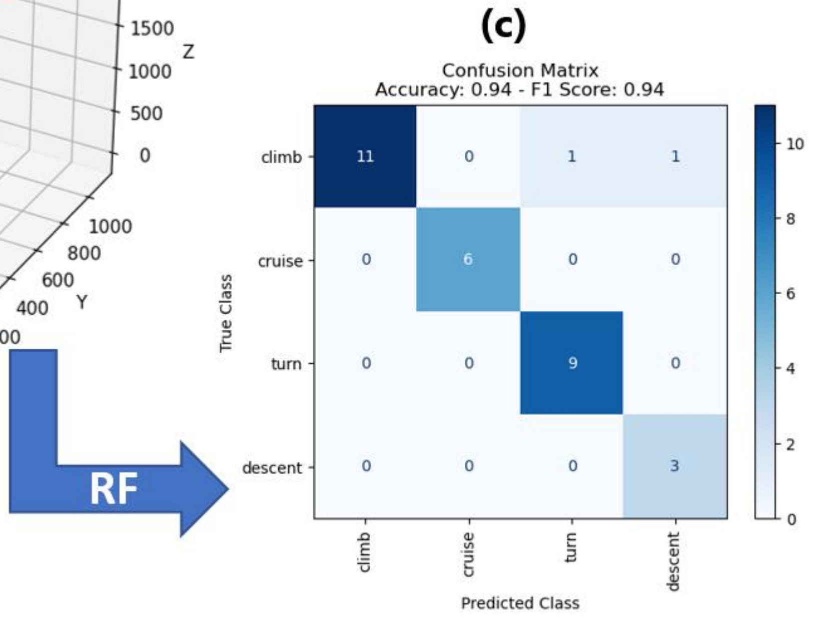
(a) Schematic of the NASA Lift+Cruise reference vehicle concept[10]; (b) Lift+Cruise propulsor and control surface definitions[10]

Variable Name	Description	Notation
Pos_bii	Position of Vehicle	p_x, p_y, p_z
Vel_bii	Vehicle velocity	u, v, w
Accel_bii	Vehicle acceleration	$\dot{u}, \dot{v}, \dot{w}$
Asensed_blb	Sensed Vehicle Acceleration	$\dot{u}_{sensed}, \dot{v}_{sensed}, \dot{w}_{sensed}$
Q_i2b	Inertial to Body Attitude (Quaternion)	Q
Omeg_Blb	Angular rate (body)	p, q, r
OmegDtl_Blb	Angular Acceleration (body)	$\dot{p}, \dot{q}, \dot{r}$
Force_b	Axis aero propulsive force (body)	X, Y, Z
Moment_b	Axis aero propulsive moment (body)	L, M, N
H_b	Angular momentum (body)	h_x, h_y, h_z
Hdot_b	Angular momentum rate (body)	$\dot{h}_x, \dot{h}_y, \dot{h}_z$
Fprop_r	Forces for Rotor & Propeller	F_{1-9}
Mprop_r	Moments for Rotor & Propeller	l_{1-9}^{props}
Mmotor_r	Moments for Rotor & Propeller Motors	l_{1-9}^{motor}
H_r	Angular momentum of Rotor & Propeller	h_{1-9}^{props}
Hdot_r	Angular momentum rate of Rotor & Propeller	$\dot{h}_{1-9}^{props}$
Omega_r	Angular rate of Rotor & Propeller	ω_{1-9}
AdvanceRatio	Advance ratio of Rotor & Propeller	J_{1-9}
EngineCmd	Rotor & Propeller Command	$n_{1-9}^{command}$
EnginePwr	Rotor & Propeller Commanded Power	n_{1-9}^{power}
CtrlSurfaceCmd	Aileron, Elevator, Rudder Commands	$\delta_{a_L,a_R,e_L,e_R,r}^{command}$
CtrlSurfacePwr	Aileron, Elevator, Rudder Power	$\delta_{a_L,a_R,e_L,e_R,r}^{power}$
EngSpeed	Rotor & Propeller Rotational Speed	n_{1-9}
EngAccel	Rotor & Propeller Rotational Acceleration	$\dot{n}_{1-9}$
Failure	Flag indicating failure of Rotor & Propeller	n_{1-9}^{fail}
Time	Time of simulation	t

[10]Simmons, B., Buning, P., and Murphy, P., "Full-Envelope Aero-Propulsive Model Identification for Lift+Cruise Aircraft Using Computational Experiments," 2021.

Case Study 2: Analysis of nominal flight data



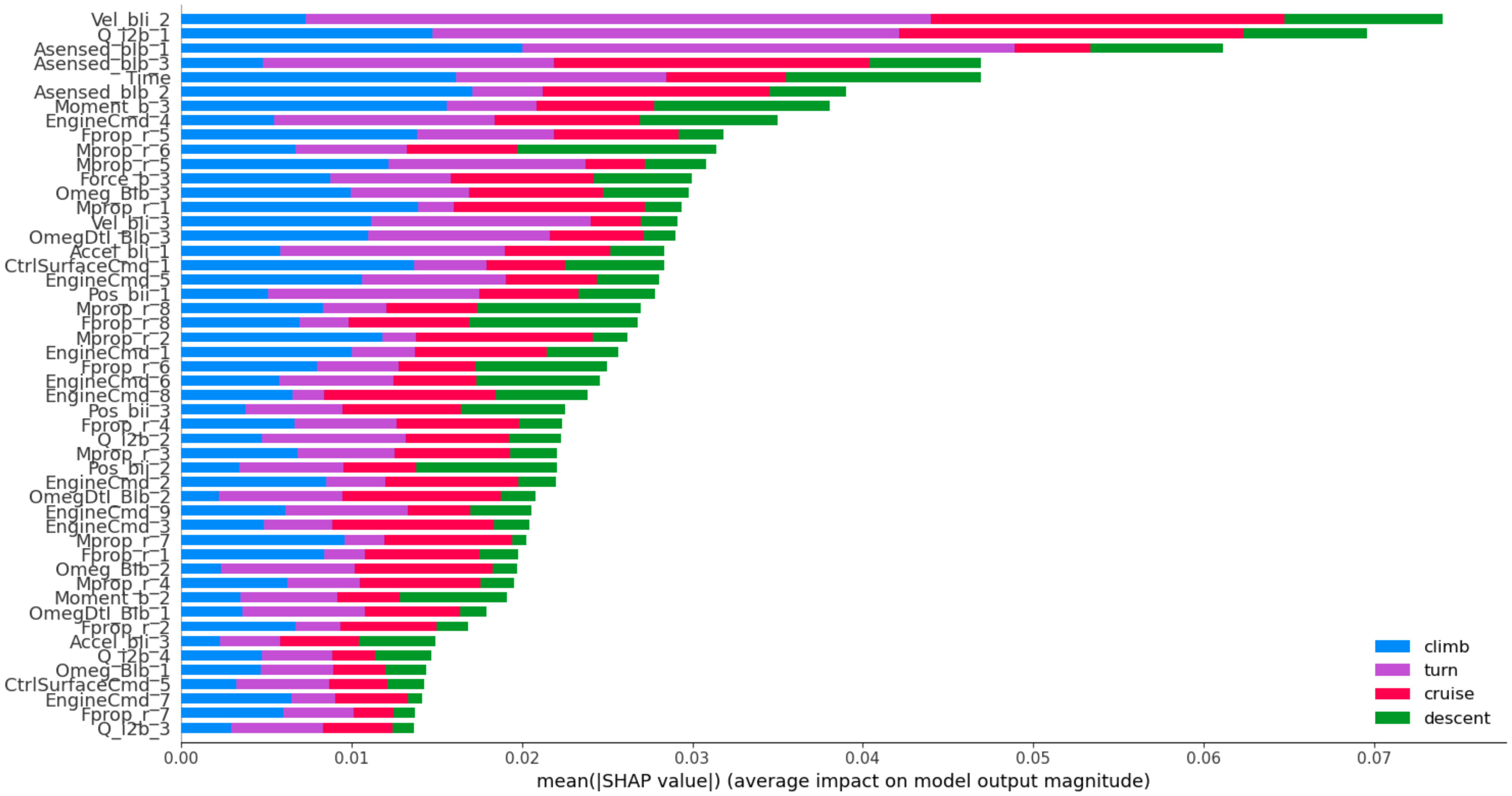


- Flight trajectory generated and labeled using Patterson et al[11]

- A Random Forest classifier was trained to predict the flight phase labels as a function of the input features

[11]Patterson, A., MacLin, G. L., Acheson, M. J., Tabasso, C., Cichella, V., and Gregory, I. M., "On Hermite Interpolation using Bernstein Polynomials for Trajectory Generation," Tech. Rep. TM-2023-0013467, National Aeronautics and Space Administration, Hampton, VA, USA, 2023.

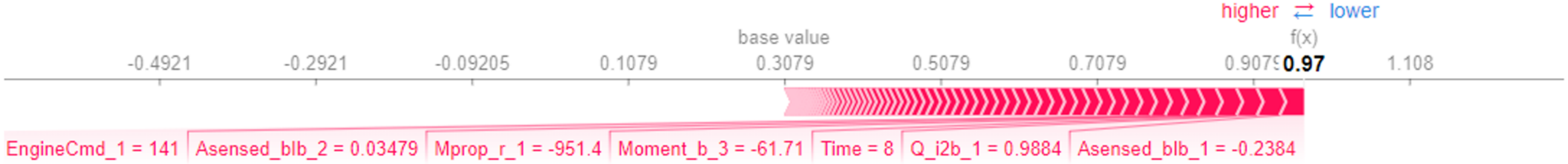
Case Study 2: Analysis of nominal flight data



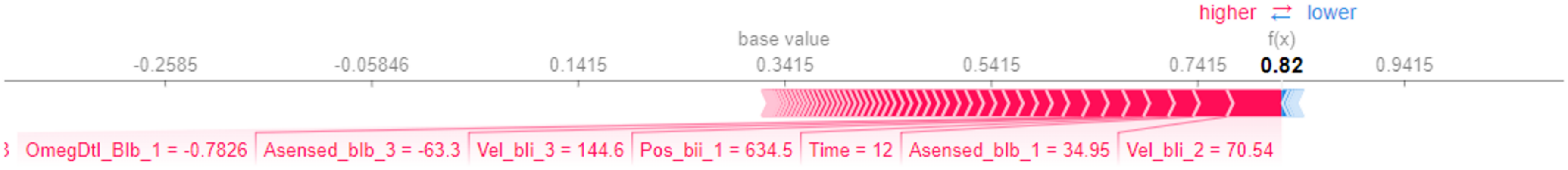
Feature importance for the entirety of the flight (all phases) based on mean absolute SHAP values

Case Study 2: Analysis of nominal flight data

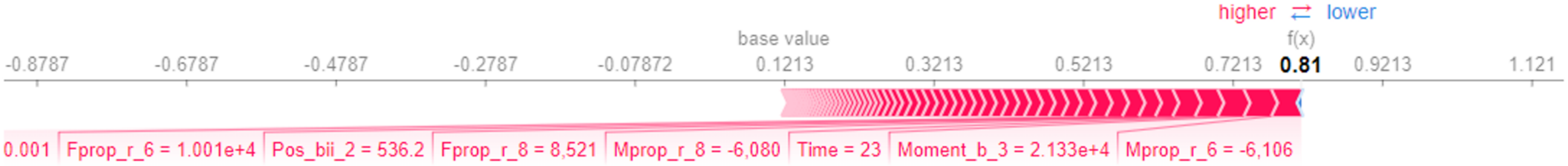
```
Enter the index of the time instance you want to plot (0 to 30 ):
8
Predicted label: climb
True label: climb
```




```
Enter the index of the time instance you want to plot (0 to 30 ):
12
Predicted label: turn
True label: turn
```

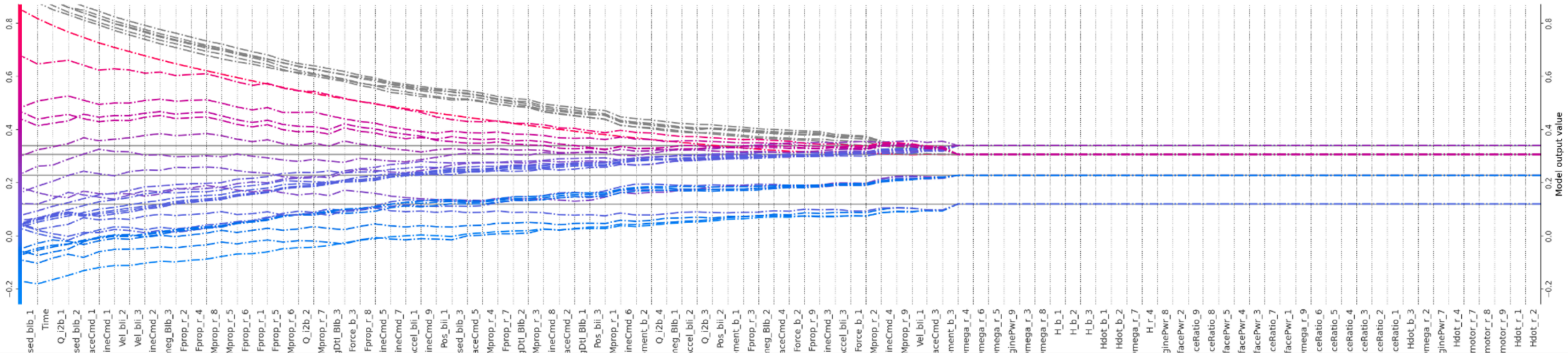


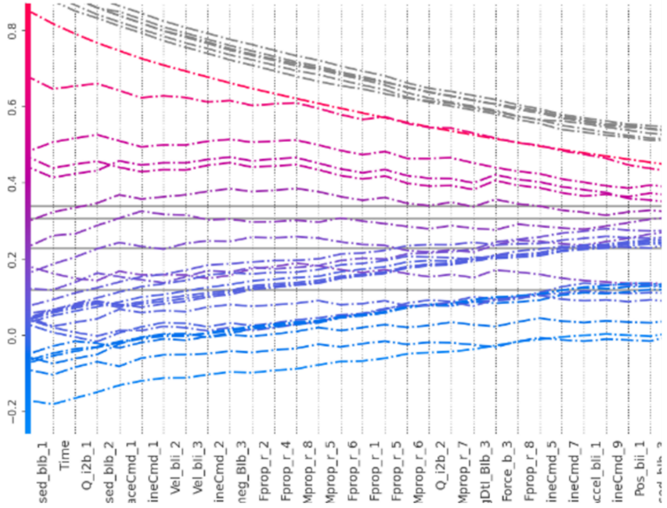

```
Enter the index of the time instance you want to plot (0 to 30 ):
23
Predicted label: descent
True label: descent
```



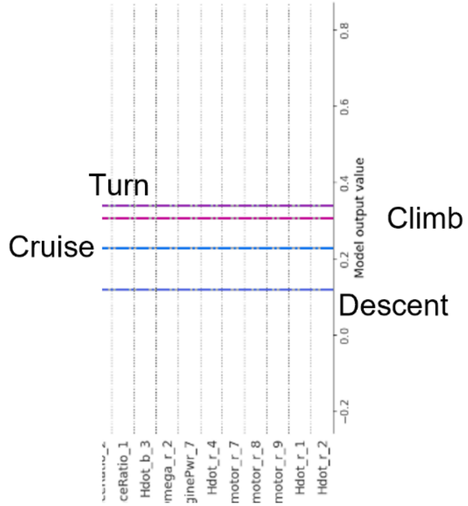
Feature importance at specific instants of time

Case Study 2: Analysis of nominal flight data





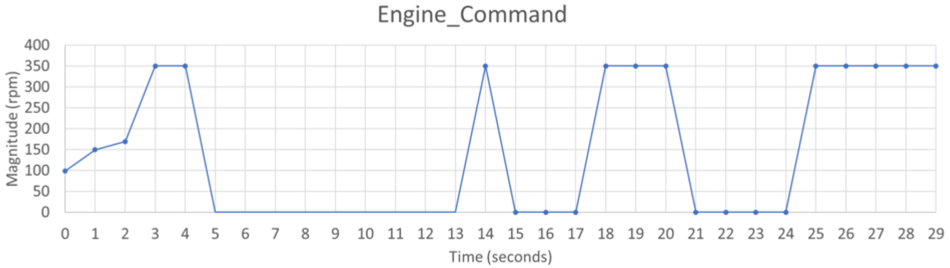
Contributing features

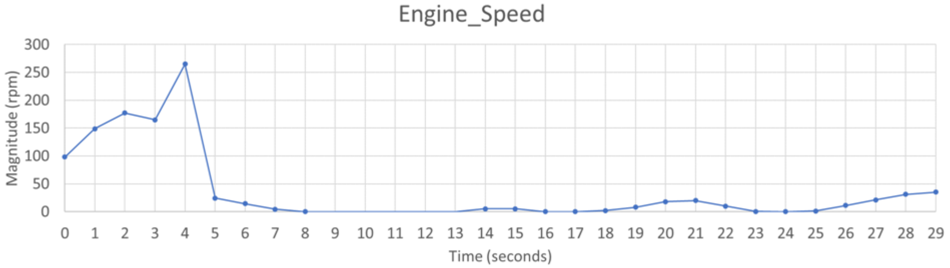


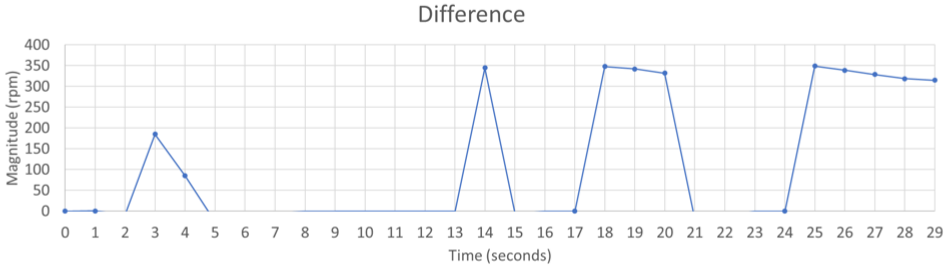
Non-contributing features

Contributing and Non-contributing features

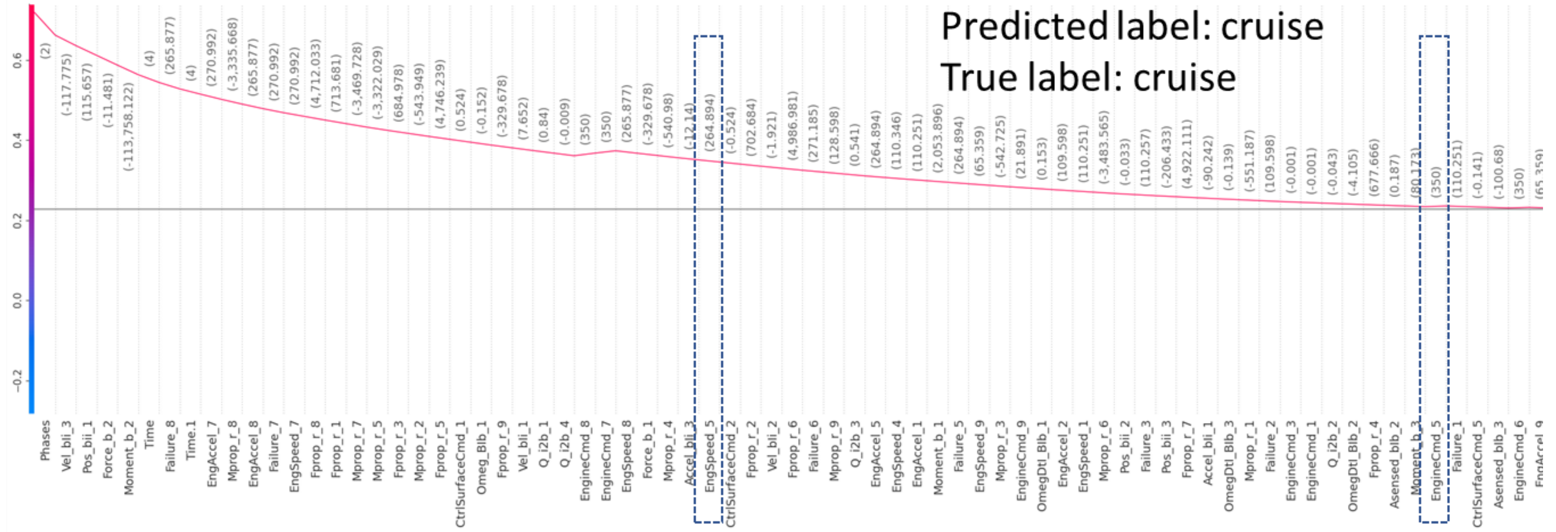
Case Study 3: Analysis of partial rotor failure

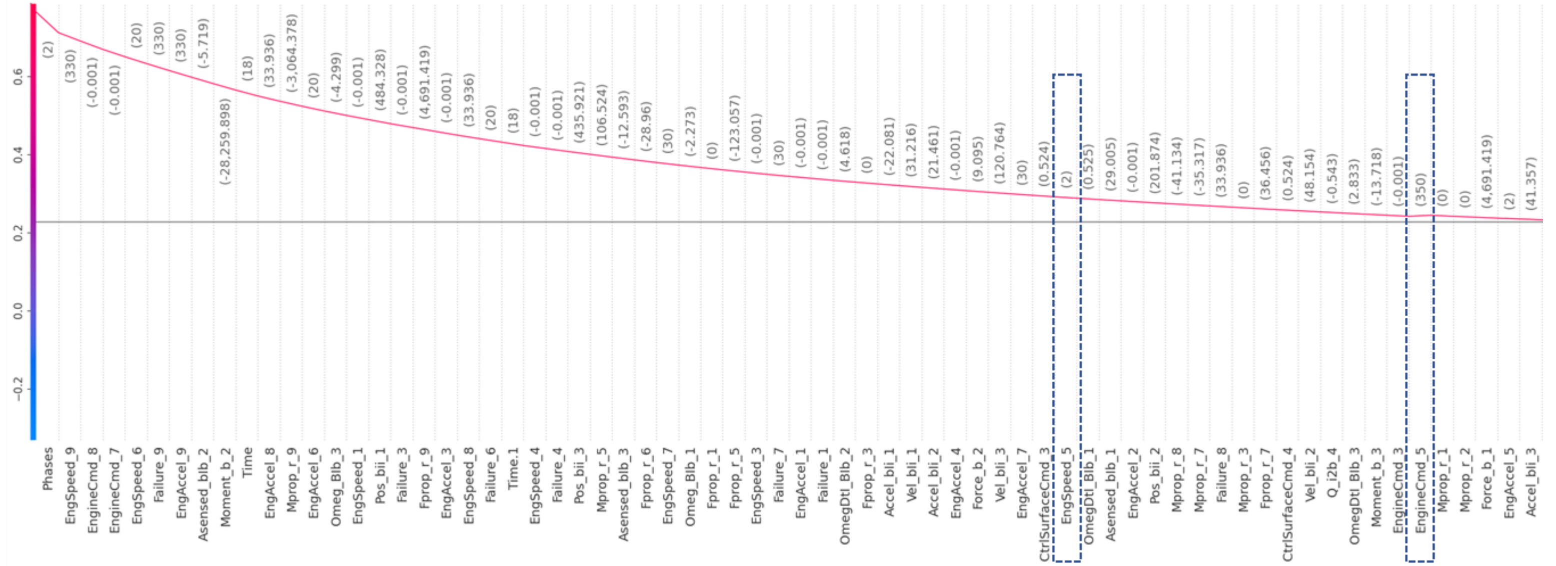






- Functionality of Propeller 5 was degraded by 90 % between time 5 s and 25 s





Shapely analysis pre (time=4 s) and post failure (time=18 s) of the rotor

Conclusion

- From a regulatory perspective, it is imperative to ensure that intricate AI models generate decisions comprehensible to humans

- This research delved into the utilization of Shapley Additive Explanations, drawing inspiration from game theory, to enhance explainability

- In the case of the simulated test function, the employed methodology successfully eliminated undesirable features that did not contribute to the output

- For nominal flight data with labels, the random forest classifier housed within the cognitive mission manager accurately predicted flight phases, and the eXplainable Artificial Intelligence (XAI) model distinctly identified both contributing and non-contributing features during that phase of flight

- Lastly, the XAI model proficiently identified the engine with partial rotor failure based on pre and post-failure Shapley analysis

- While XAI exhibits promising initial results, obstacles remain in the real-time onboard implementation of XAI, as well as with other AI models that could be time-consuming. This may give rise to challenges associated with dimensionality as the feature space expands closer to mimic the real world.

QUESTIONS ?

Contact Information:
shivakumar.ranganathan@nasa.gov
hari.s.ilangovan@nasa.gov
newtonh20@ieee.org
michael.j.acheson@nasa.gov
irene.m.gregory@nasa.gov